引用格式：孙建军, 裴雷, 马亚雪, 等. 数据驱动下智库话语逻辑及互证体系思考. 中国科学院院刊, 2022, 37(2): 153-159.
Sun J J, Pei L, Ma Y X, et al. Thoughts on data-driven discursive logic and triangulation of think tank. Bulletin of Chinese Academy of Sciences, 2022, 37(2): 153-159. (in Chinese)

数据驱动下智库话语逻辑及互证体系思考

孙建军[1,2*]　裴　雷[1,2]　马亚雪[1]　李　阳[1]
1 南京大学 信息管理学院 南京 210023
2 江苏紫金传媒智库 南京 210023

摘要　大数据资源的迅速累积与分析技术的快速发展不仅拓展智库研究范畴，促使智库研究更加重视数据洞察，同时对智库建设质量与创新水平提出更高要求。面对智库建设向现代化、创新化和科学化转变的发展需求，文章从信息链的角度阐释数据驱动智库研究变革，剖析数据变革环境下智库话语逻辑重塑的迫切需求，探讨数据驱动智库研究的多元互证体系构建方案，进而从重视相关与因果关系互补、融合技术理性与人文价值、引入三元世界分析视角、聚焦长知识链支撑、建构智库工程化服务模式等方面提出当前环境下智库建设的优化建议。

关键词　智库研究，数据驱动，话语逻辑，多元互证，信息链

DOI　10.16418/j.issn.1000-3045.20211130003

智库作为连接政策空间中多元主体的桥梁，肩负为国家提供专业化政策建议、影响政府决策的使命，亦承担引领公众了解并认知政策理念与思想的责任[1]。2016年，习近平总书记在哲学社会科学工作座谈会上指出，“智库建设要把重点放在提高研究质量、推动内容创新上”。从智库角度看，其根本就是要形成能够适应经济、社会和科技发展趋势，兼具科学创新与国际影响的智库产品与服务。面对这一目标如何引导智库向现代化、创新化和科学化的方向转变成为智库研究急需思考与解决的问题。

从信息链的角度来看，数据是智库研究的底层支撑。大数据环境与技术为智库研究提供海量数据资源与分析技术的同时，更促进数据驱动的智库研究不断发展。数据驱动下智库研究的核心是通过融合、分析与利用多源异构数据促进智库决策支持和公众引导功能的提升。在此过程中，智库话语逻辑，即如何表达与呈现研

*通信作者
资助项目：国家社会科学基金重大项目（18ZDA326）
修改稿收到日期：2022年2月8日；预出版时间：2022年2月11日

究结果，发挥着承接底层数据与智库服务的作用。换句话说，智库话语逻辑能否适应大数据环境中多元主体决策与信息需求将直接影响智库服务品质。同时，如何形成科学、稳健、客观的数据分析结论，以支撑话语逻辑建构并辅助政府决策制定亦成为决定智库研究质量的关键因素。为此，本文从数据驱动下智库研究出发，提出大数据环境下智库话语逻辑重塑及相关互证体系建设的迫切需求，旨在从数据管理与服务的视角为数据驱动下的智库研究提供优化建议。

1 数据驱动的智库研究变革

1.1 智库研究与数据分析的共生发展

智库研究与数据分析存在天然的关联。

（1）优质数据的积累是智库研究品质的重要保障。智库研究的一个重要研究路径是基于综合类数据资源、机构自建特色数据资源等来展开分析，以为决策者提供科学、可靠的决策支持。数据资源是智库研究的基础保障，数据资源及其相关要素的积累将影响智库研究的广度与深度。有效决策支持的基础是对目标问题的充分认知、分析与解读，而这一过程离不开大量数据资源的支持。换句话说，可用数据的量级、标准化、规范化、质量等在某种程度上往往能决定智库研究的发挥空间，并影响到智库分析的角度与数据分析的粒度，进而影响到智库研究产品的品质。

（2）智库研究是数据分析的重要出口。数据分析并不是搭建“空中花园”——嵌入具体场景或面向特定问题的分析才能真正发挥数据分析的潜力与价值。尤其是在大数据分析被广为推崇的新环境下，大数据分析需要依托具体场景和领域才能更好发挥针对性和有效性。智库研究通常是面向公共政策和战略问题，相关研究兼具科研属性与应用属性[2]。以智库研究为出口的数据分析，一方面能够支撑科研量化研究，使得研究结果兼具理论与实践意义；另一方面还能服务科技管理与社会活动决策，实现科研成果在现实场景中的应用。

（3）智库研究与数据分析呈现出螺旋共生关系。智库研究与数据分析并非独立存在，长期以来智库研究表现出与数据分析相互依托、共生发展的特征。智库研究为数据分析赋予现实意义，而数据分析不仅是智库研究的关键手段，更推动了智库研究的革新。大数据环境下，大量智库机构更是逐步开始主动加强数据积累，并推动自身特色数据资源建设；而数据分析方法也被智库服务大量吸收与采用，推动了数据资源的快速转化应用与价值传播。

1.2 信息链视角下的数据驱动与智库研究

信息链一般由事实（fact）、数据（data）、信息（information）、知识（knowledge）、情报（intelligence）、智慧（wisdom）等要素和环节构成；作为一个连续体，其刻画了事实被人类认知并加以利用的全过程[3]。其中，DIKW（数据—信息—知识—智慧）经典模型目前已经被学术界等广为接受，被用以揭示从数据到决策过程的演化和转化轨迹；而DIKW也是在决策分析中数据利用的一种运动过程。与之对应，潘教峰[4]提出智库DIIS三维理论模型，将智库问题研究过程抽象为收集信息（data）、揭示信息（information）、综合研判（intelligence）、形成方案（solution）4个环节。从底层逻辑来看，DIIS模型本质上与信息链对知识和智慧转化路径的解析相吻合，即智库研究也是从数据出发，通过对数据不断的加工、凝练以最终形成目标问题的解决方案。因此，可以认为信息链能够从数据层面反映智库问题的研究过程。同时，由于信息链展现了进化历程，从信息链的视角探讨数据驱动与智库研究，是一种有效揭示数据与技术环境变革对智库研究路径与模式的理论隐喻。

近年来，科学研究的数据与技术环境产生了较大程度的变革。在数据资源方面，传统研究所使用的小数据逐步被大数据甚至智慧数据取代；在信息技术方面，大数据、人工智能、云计算、物联网等新兴信息

技术不断发展，并被广泛用于解决相关研究任务。这些转变直接导致信息链中各要素实现跨层级链接，数据资源与信息技术的变革打破了传统信息链中“数据—信息—知识—智慧”的单一路径，使得数据和信息 2 类底层要素均可直接生成智慧。换句话说，通过对海量底层数据的分析即可揭示以往研究需要通过大量归纳、演绎才能获取的智慧。在此情况下，大量学者开始从底层数据出发，运用新型数据分析技术挖掘研究对象之间的关联，以期发现研究对象潜在的规律与特征，进而促成数据驱动研究的兴起[5]。相关研究使得传统研究难以量化的事实能够被大数据及其分析技术捕捉，在拓展学科研究场景与范畴的同时，实现了对传统研究模式的优化甚至颠覆。

技术与数据环境的变化同样促使智库研究从单一的专家洞察向综合性的数据洞察转变，并逐步形成数据驱动的智库研究。数据驱动的智库研究是对传统理论驱动型研究的延伸，其核心是通过运用大数据、云计算等技术发掘数据间的内在规律，进而为决策者提供决策支持[6]。① 智库研究模式方面。由数据直接产生智慧的层级跳跃有效缩减了专家在智库研究过程中的主观干预程度，这一转变从根本上使得传统智库研究高度依赖专家智慧与经验的模式发生转变。此外，数据驱动还引发了如相关思维、计算思维、工程思维、跨界思维、时空思维等一系列智库研究新思维出现，实现了对智库研究转变的进一步推动。② 智库建设方面。由于数据价值的快速提升使其成为智库的核心资源，构建完备的数据支撑体系成为国内外知名智库的核心目标。例如，美国兰德公司就面向数据驱动型智库建设目标，构建了宽领域数据体系、项目数据体系和双向综合参与型数据体系[7]。

2 数据变革环境中的智库话语逻辑重塑

2.1 智库话语体系的组成部分

智库话语体系，是智库在决策服务过程中形成的与政策话语、学术话语、媒体话语等频繁互动的表达体系。

（1）智库话语需要与政策话语相承接，智库研究需要理解政策内涵并作出针对性的回应与延展。智库研究的重要作用是辅助决策，以政策话语为基础能够更好地传达智库研究结果，明晰智库研究价值。

（2）学术话语是提升智库产品认知权威、推动国际交流的保障。可信、科学的智库研究依赖于学术研究的话语表达体系；厚重、规范的学术化研究是智库产品获得市场青睐的重要基础，也是开展开放交流的逻辑基础。相对于政策对话或企业谈判，兼容学术话语的智库话语体系将更易于跨国界传播并形成国际交流。

（3）智库话语体系需成为政府与公众之间的桥梁。智库话语不仅需要收集公众话语的共同诉求，而且往往需要向公众传播社会更易采纳的媒体话语体系。特别是在新媒体环境下，如何能够促进智库研究结果易于理解并被公众所悦纳，是提升智库的社会影响力、提升智库的实质服务质量的关键问题。

为此，本文认为数据驱动智库研究的话语体系应在政策话语与学术话语的基础上，兼容适用于与公众进行交流的媒体话语，以增强智库研究的公众传播力，并且实现对决策的影响（图 1）。

2.2 智库话语线索的关联

除了构建适应大数据环境的话语体系外，智库价

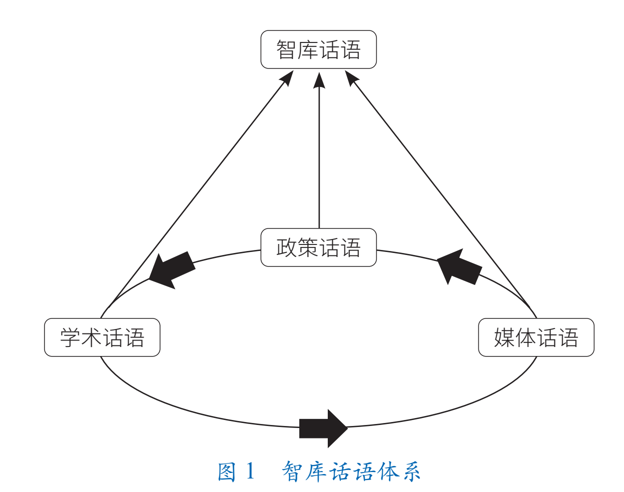


图 1　智库话语体系
Figure 1　Discourse system of think tank

值的发挥同样离不开话语线索的有效关联。换句话说，话语线索能否得到有效关联将直接影响智库研究成果的接受度及影响力。面对现代化、创新化和科学化的智库建设目标，本文认为智库话语线索的关联主要包括3个方面内容。

（1）自源话语与他源话语的融合。作为决策支持者与公众引导者，智库研究应秉承兼容并蓄的理念，审视自源话语与他源话语间的关联，并利用海量数据资源和新型分析技术对比与验证不同研究结论的置信程度，以形成科学、系统的话语线索关联。

（2）政策话语与学术话语的融合。智库研究者及其受众的认知与表达模式通常存在差异。数据驱动的智库研究在弱化专家主观涉入程度的同时，也加剧了政策话语与学术话语的关联难度。为此，大数据环境下的智库话语逻辑有必要利用大数据分析技术实现对政策情境知识与学术概念知识的关联，进而促进政策话语与学术话语的融合。

（3）政策话语与媒体话语的融合。智库不仅服务于决策者，还需要发挥引导公众的作用。媒体是实现决策与公众沟通的有效途径，特别是大数据环境下新科技、新媒体的广泛使用，更是缩短了决策者、智库与公众的距离。为此，智库研究需要有针对性地减弱话语体系不对接的问题，通过构建相关议程网络、重塑媒体话语逻辑等，以提升智库研究成果的传播能力与效果。

3 面向话语逻辑重塑的多元互证体系建设

3.1 数据驱动智库研究的困境

数据驱动从思维、技术、流程等方面都深刻影响智库研究，并促使智库研究发生巨大变革。然而，数据驱动的智库研究在具体实践中也暴露出较多的问题，使得智库研究结果的科学性和可信度有待考证。目前，数据驱动智库研究常见的困境主要包括2个方面。

（1）智库研究过度追求量化分析过程而非结果的可靠性。当前，智库研究领域广泛可见各类冠以“指数”“竞争力”“大数据分析”“百强榜”等的量化研究，此类研究旨在通过量化和数据分析来反映相关问题或现象的变化程度和发展特点。但是，一些研究没有开展系统性的前期选题论证和需求调研工作，使得研究结论浮于表面；同时，由于数据积累不足，针对同一研究问题往往生成不同的结论。这些问题的出现究其根本是由于相关研究缺乏全局视角，过分“信任”小规模或特定领域数据资源生成的结果，过于关注自身研究的“理论自洽”，而忽视从多维度对研究结论的论证，忽视对“客观真实”的深入理解，是学术话语在智库话语中的过度使用。

（2）智库研究结果存在可重复性危机。*Nature* 期刊的调查发现大量科研人员认为科学研究存在的“可重复性危机”（reproducibility crisis）[8]。同样，Serra-Garcia 和 Gneezy[9]也通过实验证实心理学和经济学研究的可重复比率较低。科研机构作为智库研究的主力，科学研究的可重复性危机也将同样影响数据驱动的智库研究结果的稳健性。而可重复性危机出现的根本原因则可能与数据自身的合理性、科学性与置信度有关[10]。

上述2类问题从根本上看均是数据本身及数据分析的不完整、不可靠、不精确等导致的连带性影响。为此，有必要引入以数据互证为基础的多元互证体系，将不同的数据与不同的方法按照一定原则进行组合和结合，以保障和提升智库研究成果的品质。

3.2 多元互证体系提升智库研究成果品质

多元互证（又称“三角互证”）来源于社会科学、教育学等领域，其强调理论、数据、方法等研究过程中核心要素的融合，而非简单的叠加。与现有很多研究不同，多元互证所秉承的融合理念是以解决同一个研究问题为目标，其中，不同要素的融合并不意味着各自失去本身的特点，而是从不同角度反映或解决目标问题。多元互证的核心思想是超越由单一要素引发的偏见，实现对研究发现可信度的提升；多元

互证主要包括数据互证（data triangulation）、研究者互证（investigator triangulation）、理论互证（theory triangulation）与方法互证（methods triangulation）4种类型[11]。

数据驱动智库研究话语逻辑的建构目标是促进政策、学术与媒体话语的融合，并面向目标受众进行有针对性的话语线索关联，进而实现智库研究价值最大化、提升智库研究影响力。然而，智库话语逻辑有效建构的前提是生成具有一致性、前瞻性和可证伪性的数据分析结果。换句话说，需要有科学化、系统化的互证体系保障智库研究结果的效度。如上所述，多元互证能够从多角度对研究结果的科学性、客观性和普适性进行验证。为此，本文认为可构建智库研究多元互证体系（图2），用于提升和保障智库研究的成果品质。

具体来说，智库研究的多元互证体系构建需以数据互证为基础，同时兼顾理论互证、方法互证和研究者互证。

（1）数据互证方面，智库研究的数据资料要重视广泛性、多元性，实施“自建数据+外源数据”模式。采用不同的数据对同一问题进行互证，能够最大化研究结论的真实性和有效性。在此过程中，数据的差异体现在来源的不同，而非生成方法的不同。换句话说，可以通过改变获取数据的时间、场景、对象等实现数据互证。对于智库研究而言，积累大量数据资源、构建相应的数据支撑体系是实现数据互证的前提和基础。

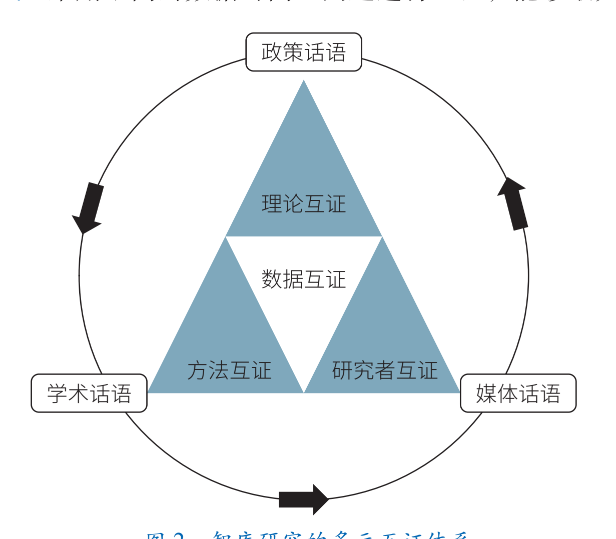


图2　智库研究的多元互证体系
Figure 2　Triangulation system of think tank

（2）理论互证方面，智库研究的数据分析和解释要以多样化的理论为基础，实现从多角度对研究结果进行诠释与解构，用以克服智库研究中的解释偏见。以相似或相异理论为视角，进行模型建构与验证分析，能够从不同视角揭示数据特征或验证分析结果的稳定性与一致性。特别是对于理论一致性较低的领域，采用理论互证能够获得更加稳健的结果。

（3）方法互证方面，量化与质性结合的混合式研究方法应该在智库研究中被推崇。方法互证包括方法内互证和方法间互证；前者是同类型方法间的相互验证，而后者则关注不同类型的方法对同一问题解析结果的差异。量化方法和质性方法的内在运行原理存在明显差异，以这2类方法为基础开展智库研究方法互证，能够有效提升研究结论的稳健性与可信度。

（4）研究者互证方面，智库研究可以邀请更多的参与者，以便从多元视角把控全局，避免问题诊断“遗漏”。通过邀请不同的观察者和访谈者参与同一研究，能够最大限度地将研究者内部的偏见最小化，进而保证研究结论的普适性。

4 数据驱动环境下智库建设思考与展望

围绕数据驱动下智库研究与建设需求，特别是对于话语逻辑及互证体系建设与优化目标，本文提出5点思考与展望。

（1）重视相关分析与因果分析融合，实现相关与因果互相补充。“因果革命”①正在席卷各个研究领

① 各领域研究者更加关注研究对象之间的因果关联，从因果而非单纯的数据角度开展研究。

域，对于智库这种智囊式研究更需对问题的内在原因及其产生机制进行深入探究。为此，可采用“相关+因果”的模式开展多元互证，以把握事物发展的内在机理，促使智库研究走向“真实世界”。

（2）注重技术理性与人文价值的融合。在新技术环境下，数据驱动的智库研究呈现出明显的技术属性。如何超越这种现代追崇的“唯技术论”，将其与社会治理的问题和需求进行高度耦合，凸显人文价值，是智库研究与服务需要思考的一个关键问题。

（3）开展三元世界下的数据互证。随着信息技术的发展，人类社会活动同时在物理空间、社会空间和信息空间（三元世界）中进行。由于数据流动的不断增强与数据量的急剧增加，物理世界、人类社会和信息空间的交互频率越发频繁；为此，不同空间中的数据也可能存在同样的研究焦点。在此基础上开展三元世界下的数据互证，将有可能产生新的洞察结果，并为智库研究提供新思路。

（4）聚焦重点科学问题的长知识链支撑。大数据时代，单一领域的知识通常难以应付复杂的科学与应用问题；为此，需要有意识地融合跨领域知识实现对智库研究更有效地支撑。具体来说，如：推动工作坊机制，构建更长的知识链支撑；推动方法池建设，完善数据池或特色数据库建设；开展系统规划与设计，落实数据体系建设的责任目标制等。

（5）建构智库工程化服务模式，打造智库产品线。数据驱动的智库研究应在多元互证体系下保障“投入—产出”的综合平衡，通过打造协同服务平台，建构集成化、协同化、系统化的工程化服务模式，以有效提升智库研究的效率与效果。

参考文献

1 潘教峰. 智库研究的双螺旋结构. 中国科学院院刊, 2020, 35(7): 907-916.

Pan J F. Double helix structure of think tank research. Bulletin of Chinese Academy of Sciences, 2020, 35(7): 907-916. (in Chinese)

2 Weaver R K. The changing world of think-tanks. PS: Political Science and Politics, 1989, 22(3): 563-578.

3 马费成, 李志元. 中国当代情报学的起源及发展. 情报学报, 2021, 40(5): 547-554.

Ma F C, Li Z Y. The origin and development of information science in China. Journal of the China Society for Scientific and Technical Information, 2021, 40(5): 547-554. (in Chinese)

4 潘教峰. 科技智库研究的DIIS理论方法. 中国科学报, 2017-01-09(07).

Pan J F. DIIS Methodology of think tank of science and technology. China Science Daily, 2017-01-09(07). (in Chinese)

5 Clauset A, Larremore D B, Sinatra R. Data-driven predictions in the science of science. Science, 2017, 355: 477-480.

6 欧阳剑, 周裕浩. 数据驱动型智库研究理念及建设路径. 智库理论与实践, 2021, 6(3): 20-27.

Ouyang J, Zhou Y H. Research on path of the construction of data-driven characteristic think tank in China. Think Tank: Theory & Practice, 2021, 6(3): 20-27. (in Chinese)

7 勇美菁, 钟永恒, 刘佳, 等. 支撑兰德公司的智库数据体系建设研究. 情报理论与实践, 2019, 42(9): 69-75.

Yong M J, Zhong Y H, Liu J, et al. Research on the construction of think tank’s data system of RAND corporation. Information Studies: Theory & Application, 2019, 42(9): 69-75. (in Chinese)

8 Baker M. 1500 scientists lift the lid on reproducibility. Nature, 2016, 533: 452-454.

9 Serra-Garcia M, Gneezy U. Nonreplicable publications are cited more than replicable ones. Science Advances, 2021, 7(21): eabd1705.

10 Miyakawa T. No raw data, no science: Another possible source of the reproducibility crisis. Molecular Brain, 2020, 13(1): 24.

11 Malamatidou S. Corpus Triangulation: Combining Data and Methods in Corpus-based Translation Studies. London: Routledge, 2017.

Thoughts on Data-driven Discursive Logic and Triangulation of Think Tank

SUN Jianjun[1,2*] PEI Lei[1,2] MA Yaxue[1] LI Yang[1]

（1 School of Information Management, Nanjing University, Nanjing 210023, China;

2 Jiangsu Zijin Media Think Tank, Nanjing 210023, China）

Abstract The rapid accumulation of data resources and the development of analysis technologies have expanded the scope of think tank research, and prompted think tanks to pay more attention to data intelligence. Meanwhile, higher requirements are put forward on the quality and innovation of the think tank. Facing the development needs of think tanks, i.e., modernization, innovation, and conscientization, this paper demonstrates the change of data-driven think tank researches from the perspective of the information chain. The paper analyzes the urgent need to reshape the discursive logic of the think tank, and discusses the construction scheme of triangulation for data-driven think tank research. Finally, several suggestions for optimizing the construction of think tanks are put forward, such as paying attention to the complementarity of correlation and causality, integrating the technical rationality and humanistic values, analyzing from the perspective of cyber-physical-human ternary space, seeking support from multidisciplinary knowledge, and constructing the engineering services of the think tank.

Keywords think tank, data-driven, discursive logic, triangulation, information chain



孙建军 南京大学信息管理学院院长、教授、博士生导师，江苏紫金传媒智库副主任。中国社会科学情报学会副理事长。主要研究领域：大数据分析、网络信息资源管理等。主持国家社会科学基金重大项目、国家自然科学基金等项目多项。E-mail: sjj@nju.edu.cn

SUN Jianjun Professor, Doctoral Supervisor, and Dean of School of Information Management, Nanjing University, and Deputy Director of Jiangsu Zijin Media Think Tank. At present, he is Vice Chairman of the Chinese Information Society of Social Sciences. His research focuses on big data analysis and online information resource management. He has obtained grants from a number of funding parties, such as the Major Program of the National Social Science Foundation of China, and the Program of National Natural Science Foundation of China. E-mail: sjj@nju.edu.cn

■责任编辑：岳凌生

*Corresponding author